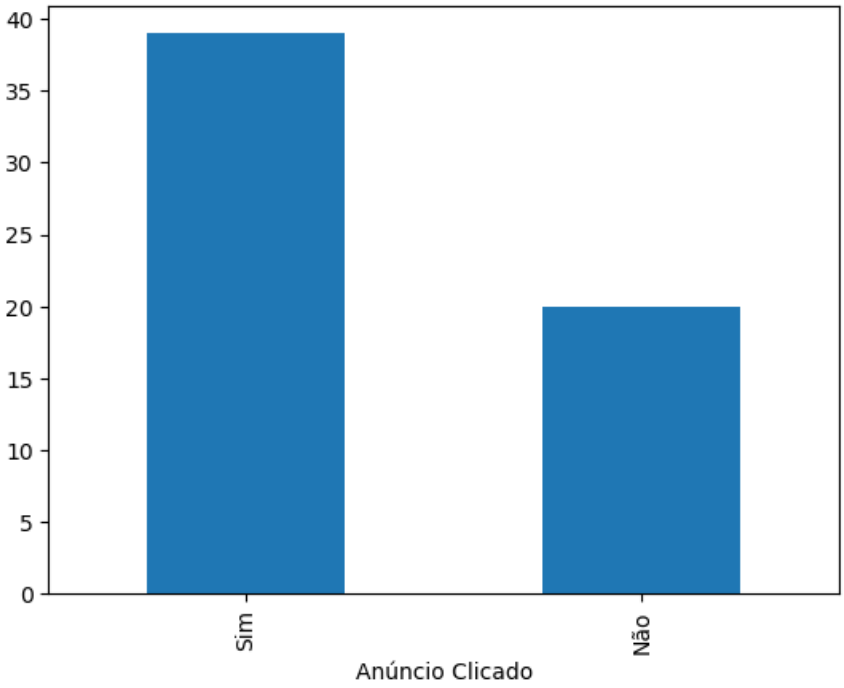
1. Análise Exploratória de Dados:
    a. Visão Geral:
    O dataset utilizado contém 6 colunas e 200 linhas. Dentre as colunas, “Idade”, “Renda Anual (em R$)”, ‘Gênero”, “Tempo no Site (min)”, “Anúncio Clicado”, “Compra (0 ou 1). Sendo esta última, a variável alvo. As variáveis foram distribuídas em palavras (object), números inteiros (int) e números decimais (float).
    b. Distribuição de Variáveis:

    De acordo com o dataset disponibilizado, conclui-se que:

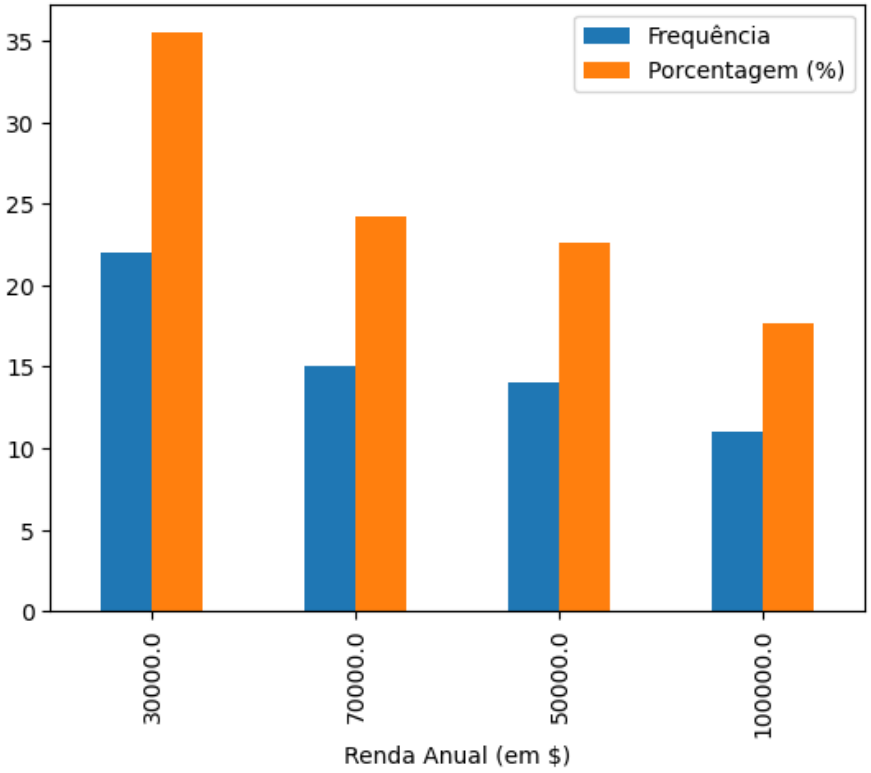
    - 66% dos clientes que compraram um imóvel, clicou em um anúncio no site;




    - 52% dos compradores são homens e 48% são mulheres;

| Gênero | Frequência | Porcentagem (%) |
|---|---|---|
| Masculino | 33 | 51.6 |
| Feminino | 31 | 48.4 |

- Compradores com renda anual média de R$30.000 representam 35.5% dos compradores gerais;




- Em média, as pessoas que gastaram mais minutos no site, compraram um imóvel.

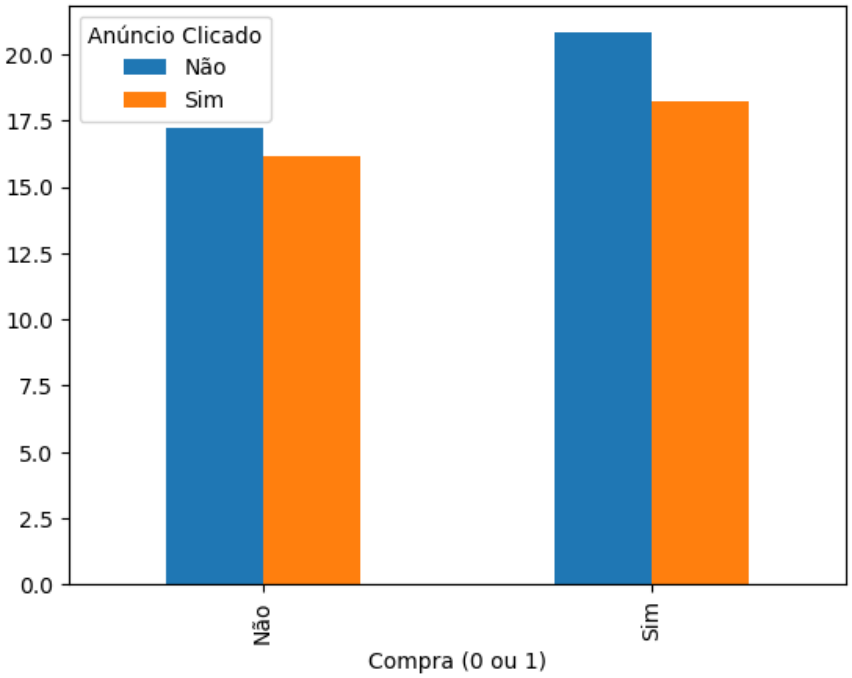



Gráfico Compra (0 ou 1) X Média de Tempo (Minutos)

c. Valores Ausentes:

Durante a Análise Exploratória de Dados, foi possível identificar valores ausentes nas variáveis “Idade”, “Renda Anual (em R$)”, “Gênero” e “Anúncio Clicado”.

2. Pré-processamento de Dados:
    a. Tratamento de Valores Ausentes:
    Para o tratamento dos valores ausentes nas colunas “Renda Anual (em R$)’ e “Idade”, foi utilizado a média dos valores para preenchê-las. Por fim, foi utilizado a função .dropna() para remover o restante das linhas com valores faltantes.
    b. Transformação de Variáveis:
    As variáveis de “Gênero” e “Anúncio Clicado” foram transformados em variáveis binárias (0 ou 1) e, as demais colunas, em números inteiros para o melhor funcionamento do modelo de classificação.
    c. Divisão dos Dados:
    Antes da divisão dos dados ser possível, os dados foram devidamente escalados com o submódulo StandardScaler. Após isso, os dados foram divididos em treino e teste com o test_size em 0.25.

3. Construção do Modelo de Classificação:
    a. Escolha do Modelo:
    O Modelo de Classificação escolhido foi a Regressão Logística por sua popularidade em resolver problemas de classificação binária e sua facilidade de interpretação.
    b. Avaliação de Desempenho:
    O modelo teve uma acurácia de 71,74% nos testes, mostrando um bom desempenho em novos dados.

4. Interpretação dos Dados:
    a. Análise de Métricas de Desempenho:
    A diferença para o treino (68,61%) indica que ele está generalizando bem. Para uma boa comparação, foi implementado o modelo Random Forest, que apresentou uma ligeira queda de desempenho com 67,39% de acurácia.
    b. Impacto das Variáveis:
    As variáveis que mais influenciaram a decisão do modelo foram "Tempo no Site (min)" - positivamente - e "Gênero" - negativamente.
5. Conclusão:
    O projeto analisou dados de clientes para identificar fatores que influenciam a compra de imóveis. Após o tratamento e organização das informações, foi treinado um modelo de regressão logística, que alcançou uma acurácia de

71,74% nos testes, mostrando bons resultados. O tempo gasto no site teve impacto positivo, enquanto o gênero influenciou negativamente. Um modelo alternativo, o Random Forest, apresentou desempenho inferior, confirmando a escolha adequada do modelo inicial.

Em resumo, o modelo forneceu informações valiosas para apoiar decisões estratégicas, com potencial para melhorias futuras.